Kickoff meeting for the proposed Provenance Working Group

Dave Dubin

April 7, 2017

1. Status of working group and interest group

1. Status of working group and interest group
2. Connection to other RDA activities and concerns

Plan for this session

1. Status of working group and interest group
2. Connection to other RDA activities and concerns
3. Work plan and deliverables

1. Status of working group and interest group
2. Connection to other RDA activities and concerns
3. Work plan and deliverables
4. Next steps and your role

1. The Scholarly Link Exchange Working Group

Some discussions this week

1. The Scholarly Link Exchange Working Group
   - Scholix: an evolving lightweight set of guidelines for exchanging information about the links between scholarly literature and data.

1. The Scholarly Link Exchange Working Group
   - Scholix: an evolving lightweight set of guidelines for exchanging information about the links between scholarly literature and data.
   - Most of the discussions on Wednesday were sense making exercises:

Some discussions this week

1. The Scholarly Link Exchange Working Group
   - Scholix: an evolving lightweight set of guidelines for exchanging information about the links between scholarly literature and data.
   - Most of the discussions on Wednesday were sense making exercises:
     - Why only publications and data sets?

Some discussions this week

1. The Scholarly Link Exchange Working Group
   - Scholix: an evolving lightweight set of guidelines for exchanging information about the links between scholarly literature and data.
   - Most of the discussions on Wednesday were sense making exercises:
     - Why only publications and data sets?
     - Hub model: why is a bridging service needed?

Some discussions this week

1. The Scholarly Link Exchange Working Group

   - Scholix: an evolving lightweight set of guidelines for exchanging information about the links between scholarly literature and data.
   - Most of the discussions on Wednesday were sense making exercises:
     - Why only publications and data sets?
     - Hub model: why is a bridging service needed?
     - Level of detail appropriate for a bibliographic record?

Some discussions this week

1. The Scholarly Link Exchange Working Group
   - Scholix: an evolving lightweight set of guidelines for exchanging information about the links between scholarly literature and data.
   - Most of the discussions on Wednesday were sense making exercises:
     - Why only publications and data sets?
     - Hub model: why is a bridging service needed?
     - Level of detail appropriate for a bibliographic record?
     - Will/can/should Scholix offer some of the services on my wish list?

Some discussions this week

1. The Scholarly Link Exchange Working Group

   - Scholix: an evolving lightweight set of guidelines for exchanging information about the links between scholarly literature and data.
   - Most of the discussions on Wednesday were sense making exercises:
     - Why only publications and data sets?
     - Hub model: why is a bridging service needed?
     - Level of detail appropriate for a bibliographic record?
     - Will/can/should Scholix offer some of the services on my wish list?

2. BoF on credit for sharing scientific data and samples

Some discussions this week

1. The Scholarly Link Exchange Working Group

   - Scholix: an evolving lightweight set of guidelines for exchanging information about the links between scholarly literature and data.
   - Most of the discussions on Wednesday were sense making exercises:
     - Why only publications and data sets?
     - Hub model: why is a bridging service needed?
     - Level of detail appropriate for a bibliographic record?
     - Will/can/should Scholix offer some of the services on my wish list?

2. BoF on credit for sharing scientific data and samples

   - Beyond the *h*-index: contextual indicators for collaborations, etc.

Some discussions this week

1. The Scholarly Link Exchange Working Group

   - Scholix: an evolving lightweight set of guidelines for exchanging information about the links between scholarly literature and data.
   - Most of the discussions on Wednesday were sense making exercises:
     - Why only publications and data sets?
     - Hub model: why is a bridging service needed?
     - Level of detail appropriate for a bibliographic record?
     - Will/can/should Scholix offer some of the services on my wish list?

2. BoF on credit for sharing scientific data and samples

   - Beyond the *h*-index: contextual indicators for collaborations, etc.
   - CoBRA guidelines for the citation of bioresources in scientific journal articles.

3. TDWG/RDA WG on Metadata standards for collection stewardship attribution

Some discussions this week

3. TDWG/RDA WG on Metadata standards for collection stewardship attribution
   - OpenRIF/VIVO-ISF Ontology: Contributor roles, activities, and relationships in clinical research

Some discussions this week

3. TDWG/RDA WG on Metadata standards for collection stewardship attribution
   - OpenRIF/VIVO-ISF Ontology: Contributor roles, activities, and relationships in clinical research
   - CRediT (Contributor Roles Taxonomy) represents the diverse roles of contributors to scientific scholarly output. The taxonomy supports multiple contributors for the same roles, with basic measures of degree of contribution.

Some discussions this week

3. TDWG/RDA WG on Metadata standards for collection stewardship attribution
   - OpenRIF/VIVO-ISF Ontology: Contributor roles, activities, and relationships in clinical research
   - CRediT (Contributor Roles Taxonomy) represents the diverse roles of contributors to scientific scholarly output. The taxonomy supports multiple contributors for the same roles, with basic measures of degree of contribution.

4. IG Data Versioning constitutive meeting

Some discussions this week

3. TDWG/RDA WG on Metadata standards for collection stewardship attribution
   - OpenRIF/VIVO-ISF Ontology: Contributor roles, activities, and relationships in clinical research
   - CRediT (Contributor Roles Taxonomy) represents the diverse roles of contributors to scientific scholarly output. The taxonomy supports multiple contributors for the same roles, with basic measures of degree of contribution.

4. IG Data Versioning constitutive meeting
   - Data Citation of Evolving Data: Recommendations of the Working Group on Data Citation (WGDC)

Some discussions this week

3. TDWG/RDA WG on Metadata standards for collection stewardship attribution
   - OpenRIF/VIVO-ISF Ontology: Contributor roles, activities, and relationships in clinical research
   - CRediT (Contributor Roles Taxonomy) represents the diverse roles of contributors to scientific scholarly output. The taxonomy supports multiple contributors for the same roles, with basic measures of degree of contribution.

4. IG Data Versioning constitutive meeting
   - Data Citation of Evolving Data: Recommendations of the Working Group on Data Citation (WGDC)
   - The WGDC recommendations enable researchers and data centers to identify and cite data used in experiments and studies. Instead of providing static data exports or textual descriptions of data subsets, we support a dynamic, query centric view of data sets.

- IVOA Provenance Data Model

- IVOA Provenance Data Model
- ElabFTW open source lab notebook

- IVOA Provenance Data Model
- ElabFTW open source lab notebook
- FairDom data management platform

- Helping the RDA community adopt existing provenance management practice.

- Helping the RDA community adopt existing provenance management practice.
- We propose a working group on *provenance patterns*.

WG Value proposition

- Helping the RDA community adopt existing provenance management practice.
- We propose a working group on *provenance patterns*.
- The patterns should relate to core RDA interests, perhaps data/data and data/people relationships.

WG Value proposition

- Helping the RDA community adopt existing provenance management practice.
- We propose a working group on *provenance patterns*.
- The patterns should relate to core RDA interests, perhaps data/data and data/people relationships.
- Provenance vocabularies offer a level of generality/specificity that address what we perceive to be implementation gaps.

WG Value proposition

- Helping the RDA community adopt existing provenance management practice.
- We propose a working group on *provenance patterns*.
- The patterns should relate to core RDA interests, perhaps data/data and data/people relationships.
- Provenance vocabularies offer a level of generality/specificity that address what we perceive to be implementation gaps.
- Our goal: constructive engagement with and response to published RDA recommendations.

Medium-term Deliverables (M12)

1. A provenance use case recording system.

1. A provenance use case recording system.
2. An initial collection of provenance use cases, elicited from other interest groups and working groups.

Medium-term Deliverables (M12)

1. A provenance use case recording system.
2. An initial collection of provenance use cases, elicited from other interest groups and working groups.
3. First documented provenance design patterns generalised from use cases.

Medium-term Deliverables (M12)

1. A provenance use case recording system.
2. An initial collection of provenance use cases, elicited from other interest groups and working groups.
3. First documented provenance design patterns generalised from use cases.
4. A report on investigation of provenance sharing implementations.

Medium-term Deliverables (M12)

1. A provenance use case recording system.
2. An initial collection of provenance use cases, elicited from other interest groups and working groups.
3. First documented provenance design patterns generalised from use cases.
4. A report on investigation of provenance sharing implementations.
5. A review of existing enterprise provenance management implementations.

Medium-term Deliverables (M12)

1. A provenance use case recording system.
2. An initial collection of provenance use cases, elicited from other interest groups and working groups.
3. First documented provenance design patterns generalised from use cases.
4. A report on investigation of provenance sharing implementations.
5. A review of existing enterprise provenance management implementations.
6. A listings of provenance tools compiled from interviews with RDA members and the provenance research community.

Medium-term Deliverables (M12)

1. A provenance use case recording system.
2. An initial collection of provenance use cases, elicited from other interest groups and working groups.
3. First documented provenance design patterns generalised from use cases.
4. A report on investigation of provenance sharing implementations.
5. A review of existing enterprise provenance management implementations.
6. A listings of provenance tools compiled from interviews with RDA members and the provenance research community.
7. A directory of open and non-open provenance data collections.

8. A taxonomy for provenance use cases.

8. A taxonomy for provenance use cases.
9. Recommendations for aligning new use cases with provenance design patterns.

8. A taxonomy for provenance use cases.
9. Recommendations for aligning new use cases with provenance design patterns.
10. Lessons for provenance sharing and enterprise management implementations.

8. A taxonomy for provenance use cases.
9. Recommendations for aligning new use cases with provenance design patterns.
10. Lessons for provenance sharing and enterprise management implementations.
11. A synthesis and critical comparison of community recommendations for provenance tool custodianship.

8. A taxonomy for provenance use cases.
9. Recommendations for aligning new use cases with provenance design patterns.
10. Lessons for provenance sharing and enterprise management implementations.
11. A synthesis and critical comparison of community recommendations for provenance tool custodianship.
12. A summary of best practice principles for provenance data collection stewardship.

13. A sustainability plan for ongoing tool and data collection custodianship.